

# **The Standard Orthography of the Tetum Language**

## **115 Years in the Making**

***Issued by the Directorate of the National Institute of  
Linguistics (INL), Democratic Republic of Timor-Leste,  
August, 2004***

Determining the standard orthography of any language is a difficult and delicate operation requiring a high degree of scientific expertise: no standard spelling system has ever been devised by an assembly of native speakers lacking a professional knowledge of linguistics, however well-educated they might be in other fields. Language planning is an exercise in both engineering and architecture, no less than building a bridge or a skyscraper. Just as no-one would want to walk across a bridge or live in an apartment block designed by a committee of future users of the bridge or residents of the building untrained in engineering or architecture, no responsible government would entrust the task of standardizing its official language to anyone but professional linguists with specific expertise in the area of Timorese languages.

Looking around the world, one finds that the standard spelling systems of official languages have come into being in one of two possible ways. The older established languages (like Latin, English, Arabic, Sanskrit, Chinese) have orthographies inherited from a longstanding literary tradition determined by writers and/or scholars. Languages which were promoted to official use at a particular time, in colonial or post-colonial situations (e.g. Malay-Indonesian, Tagalog, Vietnamese, Fijian, Samoan, Maori etc.), were codified by scholars, either individual linguists and lexicographers or committees of professional linguists appointed by a national body or government. When its sovereignty was restored in 2002, the Democratic Republic of Timor-Leste had no intention of being an exception to this general rule.

The Instituto Nacional de Linguística (INL) was established in 2001, receiving its mission from the National Council of Timorese Resistance (CNRT) to plan a *unitary* East Timorese orthography, so that all the national languages could be written according to the same conventions. This commission, confirmed by the restored East Timorese government the following year, did not imply that Tetum had no existing orthography, but rather recognized the fact that the existing systems were imperfect and inconsistent, due to their having been devised by writers most of whom had no grounding or training in linguistic science.

The reasons for the defects in the popular orthographies are not difficult to find. Persons qualified to deal competently with questions of East Timorese and Tetum orthography need a wide range of qualifications, not excluding:

- University degrees or professional experience in the disciplines of phonology, historical linguistics, Austronesian linguistics, Papuan linguistics and Romance linguistics.
- A sound knowledge of Tetum, Portuguese and Malay.
- A good working knowledge of the 15 other national languages of East Timor.
- Familiarity with all the published lexicographical works on Tetum.
- Familiarity with the scientific literature on Tetum published in Portuguese, English, Indonesian and other languages.
- A clear understanding of the difference between phonetic and phonemic systems of spelling fundamental to any standardization process.
- A clear understanding of the differences between the acrolectal, mesolectal and basilectal varieties of Tetum and their relevance to phonemic spelling.

The task of INL was nevertheless not to abolish and replace the existing orthographies, but rather to correct, refine and unite them into a single, consistent standard constructed on rigorous scientific principles, yet with sufficient flexibility to be able, when appropriate, to harmonize rational orthography with compatible traditional conventions.

Since the background of Tetum orthography is therefore not revolutionary but *evolutionary*, the purpose of the present study is to give a historical overview of the gradual improvements in the writing of the language that have culminated in the present reform.

## ***The History of Tetum Orthography***

The standard Tetum orthography given force of law by Government Decree 1/2004 of 14 April 2004 was not the work of any one linguist,

but the culmination of an experimental process spanning 115 years since the publication of the first Tetum dictionary in 1889. There were eight seminal contributions to this orthographical evolution, the first four from individuals (all of them Portuguese) and the last four from committees (two of them Timorese, the other two Timorese and international):

- a) **Sebastião Maria Aparício da SILVA** (1889)  
*Diccionario de Português-Tétum*
- b) **Raphael das DORES** (1907)  
*Diccionario Teto-Português*
- c) **Manuel Patrício MENDES and Manuel Mendes LARANJEIRA** (1935)  
*Dicionário Tétum-Português*
- d) **Artur Basílio de SÁ** (1952)  
*Notas sobre linguística timorense: Sistema de representação fonética.*
- e) **FRETILIN Literacy Committee** (1975)  
*Como vamos alfabetizar o nosso povo Mau Bere de Timor Leste*
- f) **Diocese of Dili Liturgical Commission** (1980)<sup>1</sup>  
*Ordinário da Missa: Texto Oficial Tétum and Lectionaries*
- g) **International Committee for the Development of East Timorese Languages (IACDETL)** (1996)  
*Princípios de Ortografia Tétum: Sistema Fonémico; Standard Tetum-English Dictionary*
- h) **Instituto Nacional de Linguística (INL)** (2002)  
*Matadalan Ortográfiku ba Tetun Nasionál; Hakerek Tetun Tuir Banati*

All Tetum writers of this period conformed to the orthographical conventions of one or more of these contributors and did not introduce any innovations that would find a permanent place in the official orthography of 2004.<sup>2</sup>

---

<sup>1</sup> The members were Mgr Martinho da Costa Lopes (Apostolic Administrator) and the priests António Maia, Agostinho da Costa, Francisco Tavares dos Reis, Mariano Soares, Alberto Ricardo da Silva [today bishop of Dili], Domingos Morato da Cunha, Luís Sarmiento da Costa and Leão da Costa.

<sup>2</sup> Thus Lencastre (1929) and Martinho (1942) generally followed the orthography of Silva (1887); Fernandes (1962) adhered to the orthography of Mendes and Laranjeira (1936); most émigré writers in Portugal during the Indonesian Occupation (1975-1999) conformed to the Fretilin orthography; while the *Tetun*

The division between the two kinds of contributions is historically significant, in that the four individual contributions belong to the era of Portuguese colonialism, when the only medium of instruction allowed in East Timorese schools was Portuguese, and the objective of the orthographists was to devise a *phonetic* spelling system to help foreigners learn Tetum, *not* to provide the Timorese with an indigenous literary medium. In this sense Portuguese colonialism (of the assimilationist type) was radically different from Dutch colonialism which was integrationist and encouraged the standardization of Malay as a vehicular language for the whole of Indonesia.

Consequently, attempts to overhaul the existing orthographies on an inclusivist, *phonemic* basis and to make Tetum a national and official language came only in the era of decolonization, beginning with the literacy campaign of the short-lived Fretilin government of 1975, courageously continued by the Catholic clergy, revived by a group of Timorese and international linguists based in Australia and Portugal during the Indonesian Occupation (1975-1999) and completed between 2001 and 2002 by the national language authority of a newly-liberated East Timor.

## ***The Phonemes of Tetum***

Since a scientifically valid orthography of any language requires adequate symbols for the phonemes of that language, a presentation of the inventory of Tetum phonemes is a prerequisite for any discussion of their representation in writing. The phonemic symbols given between slashes below conform (apart from the nasal vowels) to the graphemes now adopted in the standard orthography.

Even after the discovery of the phoneme by the British linguist Daniel Jones in the early 20<sup>th</sup> century, no attempt was made in Timor to devise a phonemic orthography for Tetum. The emphasis was rather on accurately and consistently representing the sounds of the language, without any attempt to analyse their mutual relationships.

---

*English Dictionary* published in by Cliff Morris in Australia in 1983 (an unacknowledged English translation and edition of the 1936 Mendes-Laranjeira dictionary) differed from the original work by jettisoning accents and introducing the phonemic graphemes <k> and <s>. Catholic writers in Timor during the Occupation followed, often with little internal consistency, the model set in 1980 by the Diocese of Dili's Liturgical Commission in 1980. Luís Costa's *Dicionário de Tétum-Português* (2000), an acknowledged revision and amplification of the 1936 Mendes-Laranjeira dictionary, favoured the Fretilin orthographical norms.

Tetum has five oral vowels:

/i/ /e/ /a/ /o/ /u/

with five nasal counterparts:

/ĩ / /ẽ / /ã / /õ / /ũ /

Both series of vowels may be long or short when tonic, or occur in the sequences (the first vowel in each sequence being stressed):

/aa/ /ae/ /ai/ /au/ /ea/ /ee/ /ei/ /eo/ /eu/ /ia/ /ie/ /ii/ /io/ /iu/  
/oa/ /oe/ /oi/ /oo/ /ou/ /ua/ /ue/ /ui/ /uo/ /uu/

There are 21 consonants:

/p/ /b/ /v/ /m/ /w/ /t/ /d/ /s/ /z/ /n/ /r/ /rr/ /l/ /ll/ /ñ/ /x/ /j/  
/k/ /g/ /' / /h/.

Of these consonants, /w/ occurs in the rural dialects but is replaced by /b/ in the lingua franca (Tetum-Praça). Some Tetum-Praça speakers efface the glottal stop /' /.

Some of the phonemic symbols above differ from their counterparts in the International Phonetic Alphabet, namely:

/r/	[ɾ]	/x/	[ç]
/rr/	[r]	/j/	[ʒ]
/ll/	[ʎ]	/'	[ʔ]
/ñ/	[ɲ]		

## ***The Evolving Principles of Tetum Orthography***

Tetum orthography reflects its historical and cultural context: the 19<sup>th</sup>-century Portuguese literary tradition. Portugal transmitted to Tetum the Roman alphabet of 26 available letters and a series of diacritics proper to the Portuguese language: the cedilla ('little z', originally a subscript Z), the tilde (originally a superscript N), and the acute, grave and circumflex accents.

In the following overview an example in italics (*becic*) represents the spelling of a particular writer; the equivalent in boldface (**besik**) represents the modern standard spelling, and an example in italics and boldface (***besik***) represents a particular graphy that coincides with the modern standard.

## ***The Portuguese Foundation***

The first Portuguese to elaborate spelling conventions for Tetum was Father Sebastião Aparício da Silva, a Catholic missionary whose Portuguese-Tetum dictionary was published in Macao in 1889. In reducing Tetum to Latin letters, Fr Silva established a pattern for the use of the seven consonants *b, d, f, l, m, r, t* that would remain unchanged to the present day. Consequently, discussion of the orthographical representation of the corresponding phonemes /b/ /d/ /f/ /l/ /m/ /r/ /t/ will not feature in the rest of this study.<sup>3</sup>

In other respects the Silva orthography (never fully consistent) adhered closely to Portuguese conventions. In general (but with some important exceptions) Portuguese loanwords were spelt exactly as in the donor language, without any adaptation to Tetum phonetics, e.g.

*lenço* **lensu** *licença* **lisensa** *machado* **maxadu** *gentio* **jentiu**  
*general* **jenerál** *marinheiro* **mariñeiru** *junho* **Juñu** *julho* **Jullu**  
*coelho* **koellu** *chá* **xá** *chouriço* **xourisu** *martello* **martelu**

In native words the consonantal phoneme /k/ (except before *e* and *i*) and intervocalic /s/ were spelt according to Portuguese rules:

*camútis* **kamutis** *crécas* **krekas** *cúac* **kuak** *hameríc* **hamriik**  
*húcic* **husik** *fácè* **fase** *táci* **tasi** *bóçòc* **bosok** *haçára* **hasara**

<sup>3</sup> Taking as their model the Tetum dialect of the southern kingdom of Samoro, both Silva and Dore recorded numerous forms with final **-l**, which has been generally replaced by **-r** in the Dili dialect of Tetum. The following examples thus have a distinctly archaic look today: (Silva) *mámal* **mamar**, *cabúal* **kabuar**, *hadél* **hadeer**, *nacfácal* **nakfakar**, *bócal* **bokar**, *fítel* **fitar**, *dúcul* **dukur**, *dícul* **dikur**; (Dores) *mámal* **mamar**, *midal* **midar**, *dádu* **dadur**, *kabúal* **kabuar**, *dácal* **dakar**.

## ***The Emergence of Tetum-based Conventions***

In its essential inconsistency, Silva's orthography was both conservative (imitating Portuguese conventions) and innovative: improvising new conventions suggested by the intrinsic structure of Tetum whenever Portuguese conventions were found to be cumbersome or otherwise inappropriate. Fr Silva and all his successors contributed to the gradual improvement and tetumization of the Portuguese-based orthography by introducing 21 individual innovations that took root in local writing habits and were eventually integrated into the standard spelling. These pillars of the modern orthographical standard can be listed in chronological order, beginning with the innovations of Sebastião da Silva.

### **A. INNOVATIONS OF S.A. da SILVA**

1. Use of <h> to represent the aspirate consonant [h] (voiceless glottal spirant), contradicting the Portuguese convention of silent <h> as in *haver, houve, homólogo, hesitar*:

*hêmo* **hemu**   *hamútuco* **hamutuk**   *hétan* **hetan**   *cakéhè* **kakehe**  
*hahí* **hahii**   *óhim* **ohin**   *haníhã* **hanihan** (**hanehan**)   *hadômi* **hadomi**  
*húrum* **hurun**   *róhan* **rohan**   *déhan* **dehan**   *haháloc* **hahalok**

2. Use of the apostrophe to represent the glottal stop [ʔ] not found in Portuguese:<sup>4</sup>

*ná'an* **na'an**   *dí'ac* **di'ak**   *sadí'a* **sadi'a**   *fó'er* **fo'er**   *tó'o* **to'o**  
*né'e bé* **ne'ebé**   *lá'en* **la'en**   *há'u* **ha'u**   *uá'in* **wa'in**

3. Notation of native etymologically-determined (i.e. disyllabic) double vowels.<sup>5</sup> This was consistent in the cases of:

---

<sup>4</sup> The apostrophe was maintained by all future writers except Does, who replaced it (though not consistently) with an acute accent on the following vowel (e.g. *toó to'o*, *nuú nu'u* 'manner', *laú la'o*, *lalián lali'an*, *hakruúko hakru'uk*), and Catholic translators of the Biblical texts in the 1990s (and in contradiction to the convention set by the Dili Liturgical Commission of 1980 which regularly marked the glottal stop with an apostrophe).

<sup>5</sup> As samples of etymologies, cf. *boot* < \**boat* < Old Timorese **berat**, *foos* < \**foas* < OT **beras**, *bee* < *wee* < OT \***wahir**, *teen* < \**ta'in* < OT \***taqi-ne**. The double vowels are still pronounced as such in the eastern varieties of Tetun-Terik as well as in

*kée* **kee** *faac* **faak** *bóoc* **book** *bóote* **boot** *clóote* **kloot**  
*íis* **iis** *fúuc* **fuuk** *míi* **mii** *môon* **moon** *húur* **huur**  
*bíite* **biit** *cúus* **kuus** *tíi* **tii** *sín* **siin** *léete* **leet**  
*clóoc* **klook** *hakée* **hakee**

Double vowels were written as single vowels in the cases of:<sup>6</sup>

*ás* **aas** *hás* **haas** *fós* **foos** *háte* **haat** *bé* **bee** *fén* **feen**  
*cmán* **kmaan** *tém* **teen** *tur* **tuur** *dadél* **dadeer** *haré* **haree**

There were hesitations in the cases of:

*dóoc* ~ *dóc* **dook** *rii* ~ *rím* **riin**

4. Use of postposed <n> to indicate vowel nasality in native words. This was consistent in the cases of:

*áten* **aten** *láran* **laran** *déhan* **dehan** *ícin* **isin** *lólon* **lolon**  
*fúrin* **furin** *lúbun* **lubun** *balún* **balion** *máun* **maun**

In other cases the Portuguese tilde or final *-m* were used to mark nasality:

---

Tetum-Praça. Dutch scholars of Western Tetum noted single vowels in these cases (e.g. *as*, *hat*, *let*, *tos*, *haré*, *hamrók*) because the originally long double vowels had undergone simplification in the Belunese dialects (including those of East Timor). That Tetum native speakers intuit the bisyllabic nature of these double vowels is clear from hypercorrective misspellings made by people whose idiolect does not include the glottal stop, e.g. *bo'ot* **boot**, *a'as* **aas**, *le'et* **leet**, *hare'e* **haree**, *hamro'ok* **hamrook**.

<sup>6</sup> The inconsistent notation of double vowels would be a persistent flaw of orthographies devised by non-linguists. Examples of such inconsistencies in the works of subsequent Tetum writers include: Doris (1907): *aáte* **aat**, *biite* **biit**, *huúr* **huur**, *huú* **huu**, *iis*, *leéte* **leet**, *loók* **look**, *luúto* **luut**, *mii*, *úuto* **uut** ~ *hás* **haas**, *hate* **haat**, *bote* **boot**, *bé* **bee**, *den* **deen**, *dil* **diir**, *fen* **feen**, *lan* **laan**, *lir* **liir**, *lós* **loos**, *tós* **toos**, *nu* **nuu**, *tós* **toos**; Mendes and Laranjeira (1935): *ás*, *áat*, *háas*, *háat*, *cmâan*, *cnaar*, *féen*, *béen*, *têen*, *néen*, *léet*, *sín*, *lín*, *bóoc*, *dóoc*, *bóot*, *clóot*, *núu*, *túu*, *úut*, *natóon*, *hamróoc*, *nonóoc*, *tatís* ~ *bé* **bee**, *lés* **lees**, *fós* **foos**, *tós* **toos**, *bí* **bii**, *clór* **kloor**, *lós* **loos**, *túr* **tuur**, *dadér* **dadeer**; and hesitations in the cases of: *táa* ~ *tá* **taa**, *íis* ~ *ís* **iis**, *kée* ~ *ké* **kee**, *lóoc* ~ *lóc* **look**, *moo* ~ *mós* **moo-s**, *rín* ~ *rin* **riin**, *sóor* ~ *sór* **soor**; Martinho (1943) *haat*, *têen*, *neen*, *boot*, *dóoc*, *fúuc*, *nuu* ~ *lés* **lees**, *dadél* **dadeer**, *bé* **bee**, *hatós* **hatoos**, *tur* **tuur**; Sá (1952): *têen* **teen**, *luun* **luun** ~ *namrés* **namrees**, *haré* **haree**; Diocese of Dili Liturgical Commission (1980): *leet*, *boot*, *aat*, *knaar*, *kbiit* ~ *los* **loos**, *ás* **aas**, *lolós* **loloos**, *hamós* **hamoos**, *hahí* **hahii**, *haré* **haree**; Costa (2000) *deer*, *deet*, *dook*, *faak*, *faat*, *fuuk*, *haat*, *neen*, *saar*, *taan* ~ *bé* **bee**, *bar* **baar**, *fen* **feen**, *has* **haas**, *haré* **haree**, *harís* **hariis**, *kes* **kees**, *let* **leet**, *los* **loos**, *nu* **nuu**, *ran* **raan**, *rin* **riin**, *sin* **siin**, *ten* **teen**, *tos* **toos**.



*haníhã hanehan saiã saián aihã ai-han*

*calém kaleen tém teen óhim ohin lílim lilin kídum kidun*  
*úcum ukun íbum ibun dadáum daudaun<sup>7</sup>*

5. Use of <k> to represent the voiceless velar stop [k] before e and i, instead of Portuguese *qu*:

*kiláte kilat hakérèc hakerek kíkite kikit kélen kelen*  
*cakéhè kakehe múkite mukit lókè loke makíli makili*

## B. INNOVATIONS OF R. das DORES

6. General use of <k> to represent the voiceless velar stop [k], replacing Portuguese *c*:<sup>8</sup>

*kánek kanek kakórok kakorok késsar kesar kbíite kbiit*  
*kiak kiak klétak kletak klóssan klosan knóruko knoruk*  
*kontra kontra kótuko kotuk kúkun kukun kúlite kulit<sup>9</sup>*

7. Introduction of the hyphen to indicate monosemantic compound words:

*ita-nia* ‘our’ *ida-neé ida-ne’e* ‘this’ *máno-ôan manu-oan* ‘chick’  
*mátan-délek matan-delek* ‘blind’ *máassin-midal masin-midar*  
‘sugar’ *máun-álin maun-alin* ‘siblings’ *tauko-laék ta’uk-laek*  
‘fearless’ *sala-máluko sala-maluk* ‘accomplice’ *hanôin-hikas*  
*hanoin-hikas* ‘regret’

<sup>7</sup> From Dorez (1907) onwards **-n** was used by all writers except Lencastre, who (following Fr Silva) vacillated between **-n** and **-m**: *néen, tahan, tulun, léten, daun, tinan ~ cabum, mutim, lorum, tanam, ulum, balum, naram, muçam*.

<sup>8</sup> The general use of **k** was maintained by Sá, the Fretilin Literacy Committee, the Dili Liturgical Commission and IACDETL; whereas Mendes and Laranjeira, Martinho and Fernandes reverted to Silva’s Portuguese-based use of *c* (but *k* instead of *qu* + *e, i*).

<sup>9</sup> The odd graphies *kúlite, kbíite, kótuko, knóruko*, with silent final vowels, were inherited from Silva and reflected the difficulty of Lusophones in dealing with words with final *-t* and *-k* (only *-r, -s, -l* and *-z* can normally occur as final vowels in Portuguese). This idiosyncrasy disappeared after Dorez.

### C. INNOVATION OF M. P. MENDES and M. LARANJEIRA

8. General use of <u> to represent final –u, replacing Portuguese <o>:<sup>10</sup>

*fáru faru hôtu hotu látu latu malu fútu futu*  
*natútu natutu táru taru têcu tekú útu utu hôpu hopu*

### D. INNOVATION OF A.B. de SÁ

9. Use of <s> to represent intervocalic native /s/, replacing Portuguese <ss>, <c> (before e,i) and <ç> (before a,o,u):

*tesi besi mesak bosok fisur tasak kadesi*

### E. INNOVATIONS OF THE FREILIN LITERACY COMMITTEE

10. Abolition of the acute and circumflex accents placed over the vowels e, o, a (é, ê; ó, ô, á, â) as in Portuguese to indicate open and close quality respectively, i.e. marking phonemes rather than allophones:<sup>11</sup>

*leten hesuk tesi kota kotu tasi manu<sup>12</sup>*

11. Abolition of the diphthong **ou** as in Portuguese indicating the close (allophonic) pronunciation of /o/, i.e. marking phonemes rather than allophones:

*moris hahoris horisehik lori foti sorin sorun hotu<sup>13</sup>*

12. Substituting **–aun** for Portuguese **–ão**:<sup>14</sup>

---

<sup>10</sup> Although Fr. Silva had used mainly **–o** in 1887, he occasionally substituted <u>, cf. *lakéru, táhu, hafúhu, hócu, tétu, nacônu, málu ~ môno monu, bôço bosu, hêmo hemu, mâno manu, hôto hotu, sélo selu, canuro kanuru, rua nulo ruanulu*.

<sup>11</sup> Previous writers had alternated the acute and circumflex accent over tonic vowels in an attempt to indicate vowel quality. The Fretilin Literacy Committee thus substituted a sound phonemic criterion for the phonetic criterion of their predecessors.

<sup>12</sup> Instead of the Lusoid phonetic graphies *léten, hêsuk, têsi, kôta, kôtu, tási, mânu*.

<sup>13</sup> Instead of the allophonic (and phonetic) graphies *mouris, hahouris, hourisehik, louri, fouti, sourin, hasouru, houtu*.

**nasaun** (< *nação*), **sabaun** (< *sabão*), **patraun** (< *patrão*), **misaun** (< *missão*), **kapitaun** (< *capitão*), **edukasaun** (< *educação*), **investigasaun** (< *investigação*)

13. Notation of /j/ from Portuguese g (+ e,i) as <j>:

**jestaun** (< *gestão*), **jeometria** (< *geometria*), **jigante** (< *gigante*), **jinástika** (< *ginástica*), **ajente** (< *agente*), **imajen** (< *imagem*)

14. Introduction of separate graphemes for seven polyvalent consonantal phonemes of Portuguese (Malay or non-Tetum indigenous) origin: /g/, /j/, /p/, /rr/, /v/, /x/, /z/:<sup>15</sup>

---

<sup>14</sup> This was a feature of acrolectal Tetum. The older (basilectal) tradition was to assimilate *-ão* to the native sequence **-án**, cf. Silva *obrigaçã, adoraçã, oraçã*; Mendes/Laranjeira: *sabân, perdân, galân, coraçân*.

<sup>15</sup> Earlier writers had failed to recognize these phonemes, mainly because until the interwar period, the vast majority of Tetum speakers assimilated foreign consonants to native ones. Thus Raphael das Dores had reflected in his orthography the (then prevailing) assimilation of Portuguese (or Malay) consonants to native ones, e.g.

/g/ > /k/ or /d/: *koiaba* **goiabas**, *doka, duka* **joga**, *fokado* **fogadu**, *dulka* **julga**,  
*dunilha* **gonilla**, *dardón* **gargón**

/j/ > /d/ or /i/: *dulka* **julga**, *dindún* **jejún**, *duis* **juis**, *dura* **jura**, *duramento*  
*juramentu*, *aiduda* **ajuda**, *iara* **jarra**

/p/ > /b/: *barassa* **prasa**, *bateka* **pateka**, *sebila* **sepilla**, *saba* **xapa**, *kombare*  
**kompadre**

/rr/ > /r/: *iara* **jarra**, *bura* **borra**, *katáro* **katarru**, *fora* **forra**, *sikôro* **sokorru**

/v/ > /b/, /w/: *biba* **viva**, *baretan* **vareta**, *uale (= wale)* **vale**

/x/ > /s/: *saba* **xapa**, *saruto* **xarutu**, *sinela* **xinela**, *sita* **xita**, *surisso* **xourisu**,  
*mansila* **maxila**

/z/ > /s/: *dussi* **dúzia**, *fussil* **fusil**

/ll/ > /l/: *sebila* **sepilla**, (Silva) *consêlo* **konsellu**

/ñ/ > /n/: *lina* **liña**, *korlina* **korliña**

The emergence of a Portuguese-speaking Timorese middle class gave rise to an ‘acrolectal’ variety of Tetum in which Portuguese consonants in loanwords continued to be pronounced as in Portuguese. Such phonemes were ‘polyvalent’ because they had popular (assimilated) and learned (unassimilated) allophones.

*gaveta gargón Bagia janela Jaku tijolu papa pombu  
borraxe terrenu vidru Vikeke livru xinelu xikra  
lixu zona kazaku ezami ezame*<sup>16</sup>

## F. INNOVATIONS OF DIOCESE OF DILI LITURGICAL COMMISSION

15. Sole use of the acute accent to indicate irregular (i.e. non-paroxytonic stress), cf.

(paroxytonic stress: no accent)

*laran klamar Ita-Boot rohan liafuan fiar dame hamutuk  
lalehan nafatin agradece agradese nu'udar Maromak*

(oxytonic or proparoxytonic stress: acute accent)

*Páskua manán glória nebé ne'ebé hahí hahii maibé  
katólíka apóstolu família Espíritu mistériu hahú*

16. Notation of phonemic final <e> in Portuguese loanwords rather than the phonetic graphy <i> based on the mesolectal/basilectal allophone:

*agradece agradese pontífise karidade padre*<sup>17</sup>

There were occasional inconsistencies, e.g. *eskolanti eskolante*

17. Use of <w> to represent native /w/ in borrowings from *Tetun-Terik* (this phoneme having merged with /b/ in the lingua franca).<sup>18</sup>

*we wee wainhira wa'in hewai hawelok hawain hawa'in*

There were, however, inconsistencies, i.e. alternating these with

<sup>16</sup> Already in 1889 Fr Silva had occasionally recognized the phonemic character of Tetum /z/, e.g. *dezejo dezeju, roza, meza, ezemplo ezemplu, ezame, diviza, cazo kazu*.

<sup>17</sup> These replaced the phonetic allophone-based graphies *agradesi, pontífisi, karidadi, padri*.

<sup>18</sup> Fr Silva and his successors had all used *u-* (and occasionally *o-*), e.g. (Silva) *ué wee, heuuái hewai, ua'in wa'in, uéc week, haouén haween, oáni wani*.

Tetum-Praça forms with /b/, cf. *labarik lawarik*, *aban bainrua awan wainrua*.

The incorrect use of the Terik form *wainhira* in Tetum-Praça instead of the standard **bainhira** stems from this inconsistency characteristic of ecclesiastical writers.

## G. INNOVATIONS OF THE INTERNATIONAL ACADEMIC COMMITTEE FOR THE DEVELOPMENT OF EAST TIMORESE LANGUAGES

18. Introduction of the graphemes <ll> and <ñ> (keyboard-friendly substitutes of <lh> and <nh> to replace <lh> and <nh> when representing Portuguese-derived palatal phonemes:

fa<sup>h</sup>la > **falla**   konse<sup>h</sup>lu > **konsellu**   Ju<sup>h</sup>lu > **Jullu**   toa<sup>h</sup>la > **toalla**  
karti<sup>h</sup>la > **kartilla**   sepi<sup>h</sup>la > **sepilla**   ervi<sup>h</sup>la > **ervilla**

ma<sup>h</sup>na > **maña**   Ju<sup>h</sup>nu > **Juñu**   deze<sup>h</sup>nu > **dezeñu**   pe<sup>h</sup>ñór > **peñór**  
akompa<sup>h</sup>na > **akompaña**   ko<sup>h</sup>nese > **koñese**   li<sup>h</sup>na > **liña**<sup>19</sup>

---

<sup>19</sup> The Portuguese digraphs <lh> and <nh> (medieval conventions of Occitan origin) represent the palatal consonants [ʎ] and [ɲ] respectively. A major weakness of Tetum spelling systems before IACDETL began its work in the 1990s was the use of these digraphs in Tetum. They are unsuitable for writing Tetum for the reason that Tetum, unlike Portuguese, has an aspirate *h* (the voiceless glottal spirant [h]) and the grapheme <h> had already been assigned to this sound by Fr Sebastião da Silva in 1887, a convention observed by all his successors. Consequently mutually contradictory signals were given by the traditional spellings *hahú*, *bainhira*, *manha*, in which <h> represented [h] in the first two words but [j] in the second. This problem was compounded by the fact that several languages of East Timor have the consonantal sequences /lh/ and /nh/, in some cases as common phonemes, e.g. Waimaha *lhire* ‘lightweight’, *lheo* ‘arrive’, *lha’a* ‘soul’, *nhasu* ‘seethe’, *nhese* ‘equal’, *nhii* ‘stand’; Baikenu *anha* ‘child’; Galoli *sinherin* ‘those’ etc.

Given the desideratum of a pan-Timorese orthography (already planned by Fr Artur Basílio de Sá in 1954 and confirmed as a commission to INL by the CNRT in 2001), it was imperative to find new symbols to replace **lh** and **nh** of Portuguese origin. Indonesian-influenced East Timorese who suggested the introduction of Malay-Indonesian-style **ly** and **ny** failed to take into account the fact that most Tetumophones (the speakers of the mesoplectal and basilectal varieties) do *not* pronounce these consonants as in Portuguese (in which case the spellings <ly>, <ny> might be acceptable) but rather depalatalize them to [l] and [n] after i, and to [il] [in] after other vowels. Therefore most East Timorese pronounce the Portuguese words *falha*, *cartilha*, *manha*, *linha* not /falya/ /kartilya/ /manya/ /linya/ but rather /fayla/ /kartila/ /mayna/ /lina/. The Indonesian graphies (apart from being politically contentious) are therefore inadequate, since the /y/ element is either absorbed by the preceding vowel or transposed to the *front* of *l* and *n*.

19. Elimination of Portuguese silent consonants in loanwords:

**istória** (< *história*), **eransa** (< *herança*), **otél** (< *hotel*),  
**ospitál** (< *hospital*), **batizmu** (< *baptismo*), **asaun** (< *acção*),  
**exesaun** (< *excepção*), **elétriku** (< *eléctrico*), **projetu** (< *projecto*)

20. Placing of acute accent over lengthened tonic oral vowels in oxytones of Portuguese origin, but not over non-lengthened tonic nasal vowels in oxytones of Portuguese origin,

**xá** (< *chá*)    **pás** (< *paz*)    **jís** (< *giz*)    **krús** (< *cruz*)  
**pár** (< *par*)    **súl** (< *sul*)<sup>20</sup>

---

The only solution to this problem was to introduce a diacritic to distinguish / and *n* as independent polyvalent phonemes. The initial IACDETL proposal, a macron over *L* and *n*, was logical, and a workable symbol. In the graphies *faĹa*, *kartiĹa*, *maĹa*, *liĹa* the macron gave the messages (1) “after *a*, *e*, *o*, *u* read me as /y/ before or after the consonant, according to your normal pronunciation”; (2) after *i*, ignore me or read me as a /y/ after the consonant, according to your normal pronunciation.” The problem was that these symbols were not keyboard-friendly. Modern computer keyboards allowed only *n* to be modified only with a superscript tilde, whereas / could not be modified at all.

The solution adopted by IACDETL and approved by INL in 2002 was to convert the macron into a tilde in the case of <ñ>, and into an extra <ll> in the case of <Ĺ>. The resultant substitutes, <ñ> and <ll> had the merit of association with the wider Portuguese and Romance linguistic tradition. The tilde was used in Portuguese (though placed over vowels, not *n*), and the graphy <ñ> was the counterpart of <nh> in Galician, one of the dialects of the original Galaeco-Portuguese language from which modern Portuguese evolved. As for <ll>, it was used in medieval Portuguese to represent the modern palatal before the introduction of <lh> (thereafter being restricted to representing /l/ in words which had LL in Latin: this convention, abolished in the reformed orthography of 1911 — e.g. *cavallo* > *cavalo*, *janella* > *janela*, *collegio* > *colégio* — survives in some Portuguese surnames, e.g. *Mello*, *Perestrello*, *Chrystello*). The symbol <ll> is also used in Spanish, Catalan and French, and <ñ> is shared by Spanish (the partner of Portuguese in the *Iberofonia* of today) and is used in American Usage phonetic transcription as well as being one of the symbols of the reconstructed alphabet for Proto-Austronesian and Proto Malayo-Polynesian, of which Tetum is a descendant.

In 1996 Armindo Tilman (formerly associated with the Fretilin literacy movement) suggested the convention **nn** (instead of **ñ**) to complement **ll** (e.g. *padrinnu* **padriñu**, *madrinna* **madríña**, *Espanna* **España**, *kompannia* **kompañia**). The proposal was logical, but was not imitated, probably because (unlike **ñ**) it lacked any precedent in the writing of Romance languages or in known systems of phonetic notation.

<sup>20</sup> These vowels were long rather than double (*pár* ‘peace’ [from Portuguese *par*] ~ *kabaas* ‘shoulder’ [native word < Old Timorese \***qabara**], *kór* ‘colour’ [Ptg. *cor*] ~ *door* ‘dirty’ [native]. Apparent exceptions to this rule of differentiating lusisms and native forms are the lusisms **lee** (= Ptg. *lê*), **reeve** (= *revê*), **prevee** (= *prevê*). These

**lan** (< lã)    **sin** (< sim)    **son** (< som)    **bon** (< bom)<sup>21</sup>

21. Tradition-based orthographical differentiation of homonyms in the interests of literacy. Where such words featured an etymologically-determined double vowel or long vowel, one of the homonyms was respelt with a single vowel topped with an acute accent (unless a nasal vowel or chronically atonic).<sup>22</sup>

*Graphy with double vowel*

**tuun** 'skewer'  
**aan** 'body'  
**bee** 'water'  
  
**foo** 'to stink'  
**haan** (= **ahan**) 'bean'  
**haree** 'to see'  
**huun** 'breath, spirit'  
**laan** 'sail'  
**maas** 'to yawn'  
**moos** 'clean'  
**roo** 'leaf'  
**saa** 'serpent'; 'family'  
**see** 'to turn, to present'  
**siin** 'sour'  
**taan** 'layer; fold'  
**tee** 'to defecate'

*Graphy with single vowel*

**tun** 'to descend'  
**an** 'oneself'  
**bé** 'the letter b'  
**be** 'which',<sup>23</sup>  
**fó** 'give'  
**han** 'to eat'  
**hare** 'rice plant'  
**hun** 'base; tree'  
**lan** 'wool'  
**mas** 'but'  
**mós** 'also, too'  
**ró** 'boat'  
**sá** 'what'  
**sé** 'who'  
**sin** 'yes'  
**tan** 'because'  
**té** 'the letter t'

Also orthographically distinguished were:

<b>bá</b> 'to go' (verb) 'hence' (adverb)	<b>ba</b> 'to' 'for' (prep.)
<b>sé</b> 'who'	<b>se</b> 'if'

---

are spelt with double vowels because the modern Portuguese graphies represent recent reductions of older (pre-1911) bisyllabic forms *lêe* (< Latin LE-GIT), *revêe* (< RE-VI-DET), *prevêe* (< PRÆ-VI-DET), cf. *têm* < *têem* (older spelling) < TE-NENT. By contrast in the other forms an original single tonic vowel was involved in the formation of the word, cf. *par* < PA-RE, *cruz* < CRU-CE.

<sup>21</sup> Oxytonic nasal single vowels were not lengthened in Tetum like oxytonic oral single vowels; hence such graphies as *\*lân*, *\*sín*, *\*bón* would be unnecessary and incorrect.

<sup>22</sup> The choices largely follow those of previous orthographists, cf. (Mendes and Laranjeira) *môon* 'clear' ~ *mós* 'also', *sée* 'to present' ~ *sé* 'who', *túun* 'skewer' ~ *tun* 'to descend'; **an** 'oneself', **bá** 'to go', **fó** 'to give', *hân* **han** 'food', **ró** 'boat', **tan** 'because'.

<sup>23</sup> Not written with an accent because it is always atonic.

## H. CONCLUDING REFORMS OF THE INSTITUTO NACIONAL DE LINGÜÍSTICA (INL)

Thanks to this cumulative process of evolution, by 2002, the year of our restored independence, all the essential features of the East Timorese national orthography were in place. The main challenge was to co-ordinate these conventions and to eliminate continuing inconsistencies. The principal inconsistencies were the following:

1. Failure to distinguish double and single vowels.
2. Failure to distinguish native double vowels (historically and actually bisyllabic) and Portuguese-derived lengthened tonic vowels (historically and actually monosyllabic), e.g. **kuus** ~ **krús**, **liis** ~ **jís**, **haree** ~ **kafé**, **maliboo** ~ **avó**, **hadoor** ~ **hemudór**, **harii** ~ **kolibrí**; **ai-naa** ~ **papá**.
3. Failure to mark with an apostrophe the glottal stop (pronounced in many varieties of Tetum-Praça as well as in *Tetun-Terik* dialects).
4. Confusion of /w/ and /b/.
5. Confusion of native /lh/ /nh/ with the outcomes of Portuguese <lh>, <nh>.
6. Non-use or incorrect use of the acute accent (as a marker of irregular i.e. non-paroxytonic stress).<sup>24</sup>
6. Phoneticism: a tendency to spell words phonetically rather than phonemically.
7. Macaronic spelling, i.e. spelling Portuguese loanwords in Tetum in the Portuguese manner instead of adapting them to the rules of Tetum orthography (two systems of orthography within the one language).

All these issues were satisfactorily resolved in the standard orthography for Tetum set out in two INL publications of 2002, *Matadalan Ortográfiku ba Tetun Nasionál* and *Hakerek Tetun Tuir Banati*, which brought to a conclusion the evolutionary process of Tetum orthography.

© Instituto Nacional de Linguística, 2004

---

<sup>24</sup> The logical application of this rule produced some differences in the use of the accent in Tetum and Portuguese, e.g. in the lusisms **nasionál** (< nacional), **alugér** (< aluguer), **altár** (< altar), **otél** (< hotel), **funíl** (< funil), **kapatás** (< capataz) ~ **konsul** (< cônsul), **rekomendavel** (< recomendável), **posível** (< possível). Occasional indigenizing graphies of this kind had already been used by Portuguese orthographers, e.g. **kintál** < quintal (Silva).



# ***Bibliography***

- Costa, Luís (2000). *Dicionário de Tétum-Português*. Lisbon: Colibri.
- Diocese of Dili Liturgical Commission (1980). *Ordinário da Missa: Texto Oficial Tétum* and Lectionaries. Dili.
- Dores, Raphael das (1907). *Diccionario Teto-Português*. Lisbon: Imprensa Nacional.
- Fernandes, Abílio (1937). *Método Prático para Aprender o Tétum*. Macao; rpt. Dili: Comando Territorial Independente de Timor, 1964.
- FRETILIN Literacy Committee (1975). “*Como vamos alfabetizar o nosso povo Mau Bere de Timor Leste*”. *Timor Leste*, No. 6, Suplementos 1,2,3.
- Hull, Geoffrey (1999). *Standard Tetum-English Dictionary*. Sydney: Allen & Unwin in association with the University of Western Sydney Macarthur.
- Hull, Geoffrey and Eccles, Lance (2001). *Tetum Reference Grammar*. Sydney: Sebastião Aparício da Silva Project and Instituto Nacional de Linguística.
- (2004). *Gramática da Língua Tétum*. Lisbon: LIDEL.
- Instituto Nacional de Linguística (INL) (2002). *Matadalan Ortográfiku ba Tetun Nasionál*. INL: Dili.
- Instituto Nacional de Linguística (INL) (2002). *Hakerek Tetun Tuir Banati*. INL: Dili.
- International Committee for the Development of East Timorese Languages (IACDETL) (1996). *Princípios de Ortografia Tétum: Sistema Fonémico*. Sydney: University of Western Sydney.
- Laranjeira, Manuel Mendes (1916). *Cartilha Tétum*. Dili: Imprensa Nacional; (1932) Macao.

- Lencastre, J. Garcêz de (1929). “Algumas regras gramaticais da língua tétum e vocabulário”. *Boletim da Agência Geral das Colónias*, 5/54 (Dec.), pp. 82-92.
- Martinho, José S. (1943). *Quatro Séculos de Colonização Portuguesa*. Oporto: Livraria Progredior.
- Mendes, Manuel Patrício and Laranjeira, Manuel Mendes (1935). *Dicionário Tétum-Português*. Macao: N.T. Fernandes & Filhos.
- Sá, Artur Basílio de (1952). “Notas sobre linguística timorense: Sistema de representação fonética.” *Estudos Coloniais*. Lisbon, 3 (1-2), pp. 39-60.
- Silva, Sebastião Aparício da (1887). *Diccionario de Português-Tétum*. Macao: Typographia do Seminario.
- Tilman, Armindo (1996). “*Matadalan nosi tetun — Timór lian*. Lisbon (private edition).